

ИНТЕРНЕТ-ПОРТАЛЫ И ИХ РОЛЬ В ОБРАЗОВАНИИ

УДК 378.4

DOI: 10.17223/16095944/67/4

Г.В. Можаяева, А.В. Слободская, А.В. Фещенко

Национальный исследовательский Томский государственный университет, Томск, Россия

ИНФОРМАЦИОННЫЙ ПОТЕНЦИАЛ СОЦИАЛЬНЫХ СЕТЕЙ ДЛЯ ВЫЯВЛЕНИЯ ОБРАЗОВАТЕЛЬНЫХ ПОТРЕБНОСТЕЙ ШКОЛЬНИКОВ

Представлены результаты анализа пользовательских данных из социальной сети «ВКонтакте» старшеклассников Сибирского федерального округа. Авторы оценивают возможность определения образовательных интересов старшеклассников по их профилю в социальной сети для формирования индивидуальных предложений по программам высшего профессионального образования со стороны региональных вузов.

Ключевые слова: социальные сети, абитуриенты, школьники, анализ данных, образовательный интерес.

При использовании университетом социальных сетей для рекрутинга возникают задачи, для которых нет простого и эффективного решения. Они связаны с отбором абитуриентов с сильным интересом к определенной предметной области и мотивацией к обучению. Стандартные инструменты социальных сетей по сегментированию целевой аудитории используют в основном социальные, демографические и географические данные. Для выявления потребностей и интересов абитуриентов в сфере образования этих данных недостаточно, но они могут быть дополнены информацией о пользователе, содержащейся в его профиле: подписки к тематическим группам и страницам, публикации на стене, сеть связей и т.д. Подходы, позволяющие анализировать пользовательские данные и интерпретировать их для организации эффективного информационного воздействия, уже используются в политике и маркетинге. Основу этих подходов составляют методы лингвистического анализа и психодиагностики. Но пока найденные решения не применяются университетами для выявления образовательных интересов и рекрутинга абитуриентов. Поэтому одной из задач нашего исследования является проверка предположения о возможности поиска потенциальных абитуриентов в социальных сетях для конкретных факультетов и специализаций через выявление их интересов к соответствующим предметным областям.

По нашему мнению, интерес старшеклассника к той или иной области знаний связан с вероятностью поступления на определенный

факультет университета. В социальной сети интересы пользователя проявляются через тексты, опубликованные на странице его профиля, и участие в сообществах, анализ которых позволяет определить интерес к той или иной области знаний и классифицировать всех потенциальных абитуриентов на три группы интересов: гуманитарные, естественные, физико-математические, а затем дифференцировать в каждой группе пользователей по степени проявления интереса.

Предлагаемый нами подход по анализу пользовательских данных из социальных сетей школьников и студентов можно использовать для решения разных образовательных задач: профориентационной работы со школьниками, определения психолого-педагогических особенностей для индивидуализации обучения, выявления школьников с признаками одаренности, для индивидуального сопровождения в школе и вузе, определения образовательных потребностей школьников для проектирования индивидуальных траекторий личностного и профессионального развития в рамках программ высшего образования. Основная проблема для исследований в этом направлении – отсутствие готовых методов анализа данных из социальных сетей для решения образовательных задач. Большой объем данных и их разнородность делают непригодными инструменты ручной обработки, требуют новых инструментов анализа разнородных неструктурированных данных.

Анализ пользовательских данных из социальных сетей для определения личностных харак-

теристик их владельцев является направлением в междисциплинарных исследованиях, набирающим популярность в последнее десятилетие. Большинство опубликованных работ доказывают существование связи между цифровыми данными из профиля пользователя в социальных сетях и его реальными демографическими, социальными и психологическими характеристиками. В одной из первых работ по этой теме в 2013 г. «Personality, gender, and age in the language of social media: The open-vocabulary approach» [1] авторы проанализировали огромный массив текстовых данных с целью проследить взаимосвязь между психологическими особенностями человека и используемым им языком в Facebook. Авторы признаются, что связи между языком и психосоциальными характеристиками зачастую зависят от контекста, и поэтому в другом контексте результаты могли бы быть отличны. Но тем не менее выявлена независимо от контекста сильная корреляция между тестами из профиля пользователей Facebook и их полом, возрастом, эмоциональной стабильностью, интровертностью и нейротизмом.

В работе «Manifestations of user personality in website choice and behaviour on online social networks» [2] исследуется, как личностные свойства проявляются в сети в зависимости от посещаемых ими сайтов и активности в Facebook. Для исследования был использован массив данных 350 тыс. американских пользователей Facebook. Показатели индивидуальных качеств пользователей были измерены с помощью стандартной анкеты International Personality Item Pool [3], основанной на модели Большой пятёрки. Авторам исследования удалось через анализ подписок пользователей (отметки на странице и группы «Мне нравится») с высокой точностью предсказать возраст, далее идут экстраверсия, нейротизм, добросовестность и открытость новому опыту.

Группа ученых (D. Markovikj et al.) [4] подобно предыдущей работе попыталась выяснить, возможно ли смоделировать личность пользователя на основе данных из его профиля в Facebook. В данном исследовании выборка составила только 250 пользователей (примерно 10 тыс. постов), однако количество используемых параметров существенно выше (725), чем в предыдущей работе.

В отличие от предыдущих исследований были использованы также методы лингвистического анализа данных, полученные на основе методов LIWC и NLTK¹ для работы с естественным языком. Еще один набор параметров был получен с использованием словаря AFINN², где каждому слову приписывается значение от -5 до $+5$ в зависимости от его эмоциональной окраски. Так, как и в предыдущем исследовании, экстраверсия показала высокий уровень корреляции с такими параметрами, как плотность сети друзей, количество используемых хэштегов. Значимый уровень корреляции наблюдался между добросовестностью и словами с эмоциональной оценкой $+2$.

Из обзора наиболее известных работ зарубежных исследователей по теме анализа пользовательских данных в социальных сетях следует вывод о том, что цифровые данные, оставленные человеком в сети, могут использоваться для моделирования и предсказания отдельных психологических черт личности, однако задачи по интерпретации этих данных для выявления образовательных потребностей пользователя исследователями не ставились. Данные социальных сетей в контексте образовательной деятельности исследуются в основном для выявления их маркетингового потенциала [5, 6].

Отечественным исследователем С.А. Щебенко [7] на примере пользовательских данных из российской социальной сети «ВКонтакте» также доказано существование корреляции между анализируемыми параметрами профилей пользователей с экстраверсией и открытостью опыта, которая была опосредована наличием экстраверсии в большинстве случаев. Исключением являлась лишь связь между открытостью опыта и числом собственных записей на стене. Автор предполагает, что данная черта личности может способствовать более активному поиску информации и размещению результатов такого поиска. Из всех выявленных в исследовании взаимосвязей именно последняя может позволить определить пользователей, открытых к поиску знаний, т.е. потенциально мотивированных к обучению.

Еще одна работа отечественного автора [8] посвящена возможностям лингвистики в диагностировании личности по тексту. В исследовании не задействованы данные из социальных сетей, но

¹ <http://nltk.org/>

² <http://neuro.imm.dtu.dk/wiki/AFINN>

Таблица 1

Сравнение направления подготовки студентов с тематикой сообществ

Направления подготовки	Количество проанализированных студенческих профилей	Соответствие тематики сообществ направлению подготовки		
		Соответствует на 100 %	Соответствует более чем на 30 %	Соответствует менее чем на 30 %
Гуманитарные науки	324	88 %	6 %	6 %
Физико-математические науки	199	17 %	1 %	82 %
Естественные науки	139	4 %	0 %	96 %

проанализированы результаты тестов Большой пятерки более 100 респондентов и их результаты диагностики коммуникативной установки В.В. Бойко. Для анализа было выделено 67 различных индексов, основанных на морфологии и синтаксисе. Результат исследования подтвердил, что возможность определять психологические характеристики пользователя через анализ строения предложений и выбора частей речи даёт более широкие возможности, чем обычный поиск корреляции с используемым вокабуляром [4]. Авторы работы не рассматривают возможности применения методики и повторения результатов на материалах текстов на естественном языке из социальных сетей. Но тем не менее исследование является интересным с точки зрения лингвистического анализа авторских текстов для определения особенностей образовательного потенциала у автора. Например, можно искать пользователей, тексты которых обладают высоким коэффициентом лексического разнообразия, что может указывать на их богатство речи и, следовательно, высокий уровень образованности и способность излагать свои мысли.

Таким образом, в научной литературе тема анализа пользовательских данных из социальных сетей востребована и актуальна, но не в контексте решения образовательных задач.

Данная работа посвящена проверке гипотезы о том, что пользовательские данные учащихся 11-х классов, потенциальных абитуриентов, в социальной сети «ВКонтакте» могут содержать сведения об их интересах к той или иной предметной области и указывать на будущее направление подготовки в вузе.

Цель представляемой работы – проверка выдвинутой гипотезы на одном типе открытых данных из профилей пользователей «ВКонтакте» – под-

писки на сообщества. На текущей стадии исследования проверка гипотезы ограничивается только гуманитарным направлением подготовки.

На первом этапе работы гипотеза проверялась на студентах ТГУ 1-го курса, 2016 г. поступления. В социальной сети «ВКонтакте» были найдены выпускники школ города Томска 2016 г. Выборка составила 18 000 человек. Из профилей пользователей через API «ВКонтакте» выгружены сведения о сообществах, в которых они участвуют. Из общего количества выбрано 959 наиболее популярных сообществ. Определение тематики сообщества проводилось экспертами вручную. В результате составлен классификатор сообществ и определена доля каждой тематической рубрики в общем количестве [9].

Проверка классификатора на 992 студентах ТГУ (выгрузка данных в ноябре 2016 г.) показала, что 66 % из них подписаны на группы и страницы, тематика которых может быть связана с той или иной предметной областью. Сравнение направления подготовки студентов с тематикой сообществ, на которые они подписаны, представлена в табл. 1. Нужно отметить, что в число студентов попали пользователи «ВКонтакте» из выборки 18 000. Студенты – это те же выпускники школ, поступившие в ТГУ летом 2016 г.

Точность выявления гуманитариев с помощью классификатора сообществ составила 94 %. Невысокая точность определения интересов к физико-математическому и естественнонаучному контенту можно объяснить ограниченной выборкой сообществ для составления классификатора: из 959 проанализированных сообществ 231 соответствует гуманитарной тематике, 22 – физико-математической и только 1 – естественнонаучной.

Таким образом, подтвердилась гипотеза о возможности выявления образовательных интересов

пользователя социальной сети через анализ тематических сообществ, в которых он состоит, но только на студентах гуманитарного направления подготовки.

На втором этапе исследования гипотеза проверялась на выпускниках школ 2017 г. Выборка пользователей была увеличена до 126 000 человек, в которую попали все выпускники Сибирского федерального округа, зарегистрированные «ВКонтакте». Для расширения классификатора была увеличена выборка сообществ до 75 000, в которую попали публичные страницы и группы с не менее чем 100 участниками из анализируемой целевой аудитории. Из 75 000 сообществ 1 416 оказались связанными по теме с школьными предметами: 73 % – гуманитарные, 18 % – физико-математические и 9 % – естественнонаучные (табл. 2).

Список сообществ каждого из 126 000 пользователей сравнивался с тематическим классификатором сообществ, и произведен подсчет количества сообществ всех тематик, всех образовательных сообществ, отдельно гуманитарных, естественнонаучных и физико-математических сообществ. Ранжирование всех пользователей (126 000) по отношению к количеству сообществ с гумани-

тарной тематикой позволило построить рейтинг школьников по силе выраженности интереса к гуманитарным наукам.

Для проверки предположения о «гуманитарности» школьника с высокой долей гуманитарных сообществ в подписках было отобрано 9 390 пользователей с наиболее выраженным интересом к гуманитарным предметам: доля сообществ с гуманитарной тематикой – более 75 %, количество подписок на гуманитарные сообщества – не менее 5, доля образовательных сообществ относительно всех сообществ – не менее 5 %. Каждому пользователю из выбранной группы предложено подтвердить или опровергнуть предположение о том, что он является «гуманитарием» и планирует поступать в вуз на гуманитарное направление подготовки. Опрос школьников проводился в июле 2017 г. в разгар приемной кампании в вузах и в период максимального самоопределения выпускников по поводу их дальнейшего образования.

Обратная связь была получена от 1 100 школьников, из них 904 (82 %) подтвердили предположение о гуманитарных интересах. 176 (18 %) не считают себя гуманитариями. Допуская, что самооценка школьника может быть ошибочной, «негуманитарным» пользователям предложено пройти дополнительное тестирование. Сложность выявления склонности школьника к гуманитарным, техническим или естественным наукам обусловлена тем, что большинство современных методов профессиональной диагностики не предполагают четкого разграничения между данными направлениями. Так, в одном из наиболее популярных на сегодняшний день дифференциально-диагностическом опроснике Е.А. Климова тип профессии «человек-человек» включает профессии, относящиеся как к естественному профилю подготовки (медицина), так и к гуманитарному (педагогика). Поэтому в рамках данного исследования использовался профдиагностический метод, позволяющий выявить склонность к гуманитарным наукам, а именно карта интересов А.Е. Голомштока, которая позволяет идентифицировать склонности личности в различных сферах деятельности, 11 из которых относятся к гуманитарным (история, литература, журналистика, педагогика, общественная деятельность, педагогика, юриспруденция, экономика, иностранные языки, изобразительное искусство, сценическое искусство и музыка). Для диагно-

Таблица 2

Распределение образовательных сообществ по тематике

Название предмета	Доля в общем количестве сообществ образовательной тематики, %
Иностранные языки	14
Литература	10
Психология	8
Искусство	8
История	6
Обществознание	5
Математика	4
Техника и технологии	4
Химия	3
Русский язык	3
Биология	3
Физика	3
Культура	3
Информатика	1
География	1
Философия	1
Юриспруденция	1
Геометрия	1

стики онлайн была выбрана сокращенная версия опросника из 50 вопросов.

Из 176 респондентов 94 прошли диагностический тест на определение карты интересов А.Е. Голомштока. У 54 респондентов по результатам тестирования выявлен интерес к одной или нескольким гуманитарным сферам деятельности со значением от +3 до +5 по шкале измерения от -5 до +5. Таким образом, несмотря на отрицание респондентами своей склонности к гуманитарным наукам, у них проявляется интерес к гуманитарным сферам деятельности.

В результате проведенного исследования подтверждена гипотеза о возможности определения образовательного интереса школьников к гуманитарным предметам на основе анализа подписок на тематические сообщества в социальной сети «ВКонтакте». Точность определения составляет 82%, при этом еще 5% респондентов обладают выраженным интересом не только к технической и естественнонаучной тематике, но и к гуманитарной.

Данный подход, основанный на анализе открытых больших данных, позволит университетам при организации приёмной кампании снизить затраты на маркетинг и в то же время расширить географический охват, повысить релевантность аудитории, организовать персональные коммуникации, увеличить количество высокомотивированных абитуриентов, привлечь их внимание к интересным для них направлениям подготовки. Также данный метод выявления пользователей социальных сетей с определенным образовательным интересом может быть адаптирован для продвижения различных образовательных продуктов, например массовых открытых онлайн-курсов или программ послевузовского образования.

В продолжение исследования планируется проверить гипотезу выявления образовательных интересов к естественнонаучным и техническим наукам, найти корреляцию между тематикой образовательных сообществ пользователя и программой высшего образования, на которую он поступает, разработать методику прогнозирования поступления абитуриента на конкретную программу подготовки, дополнить метод анализа подписок на сообщества лингвистическим анализом текстов из профилей пользователей.

Исследование выполнено при финансовой поддержке РФФИ в рамках научного проекта № 17-16-70004.

ЛИТЕРАТУРА

1. *Schwartz H.A. et al.* Personality, gender, and age in the language of social media: The open-vocabulary approach // *PloS one*. – 2013. – Vol. 8, № 9. – P. e73791.
2. *Kosinski M. et al.* Manifestations of user personality in website choice and behaviour on online social networks // *Machine learning*. – 2014. – Vol. 95, № 3. – P. 357–380.
3. *Goldberg L.R. et al.* The international personality item pool and the future of public-domain personality measures // *Journal of Research in personality*. – 2006. – Vol. 40, № 1. – P. 84–96.
4. *Markovikj D. et al.* Mining facebook data for predictive personality modeling // *Proceedings of the 7th international AAAI conference on Weblogs and Social Media (ICWSM 2013)*. – Boston, MA, USA, 2013.
5. *Fagerstrom A., Ghinea G.* Co-creation of value in higher education: using social network marketing in the recruitment of students // *Journal of Higher Education Policy and Management*. – 2013. – Vol. 35, № 1. – P. 45–53.
6. *Rutter R., Roper S., Lettice F.* Social media interaction, the university brand and recruitment performance // *Journal of Business Research*. – 2016. – Vol. 69, № 8. – P. 3096–3104.
7. *Щебетенко С.А.* Большая пятерка черт личности и активность пользователей в социальной сети «ВКонтакте» // *Вестник Южно-Уральского государственного университета. Серия: Психология*. – 2013. – Т. 6, № 4.
8. *Литвинова Т.А.* Возможности компьютерной лингвистики в диагностировании личности по тексту // *Вестник Воронежского государственного университета. Серия: Лингвистика и межкультурная коммуникация*. – 2015. – № 3.
9. *Feshchenko A., Goiko V., Mozhaeva G. et al.* Analysis of user profiles in social networks to search for promising entrants // *INTED2017 Proceedings, 11th International Technology, Education and Development Conference, March 6th-8th, 2017*. – Valencia, Spain. – P. 5188–5194.

Mozhaeva G.V., Slobodskaya A.V.,
Feshchenko A.V.

National research Tomsk State University,
Tomsk, Russia

INFORMATIONAL POTENTIAL OF SOCIAL NETWORKS FOR REVEALING PUPILS EDUCATIONAL NEEDS

Keywords: social networks, entrants, pupils, data analysis, educational interests.

Our experience shows that the interest of high school students in one or another subject field is connected to the probability of entering a particular university faculty. A user's interests are presented in social networks via the texts published on his/her profile wall and participation in groups. Analyzing these texts and groups, to our mind, we are able to define the research areas this user is inclined to. Then we can divide possible entrants according to their interest into three groups (Humanities, Life Science, Physics and Mathematics) and differentiate users in every group according to the extent of their interest.

This study is devoted to testing the hypothesis that pupils' user data in social networks may signal about their interests in some of the subject fields and their future direction of studying at a university. In this case, we mean pupils studying in high school who are considered to the potential entrants; the analyzed social network is "VKontakte". We aim at testing this hypothesis on one type of open data from users profiles only which is subscriptions to the communities. At the current stage of the study we work with the Humanities faculties only.

The hypothesis has been tested on the school leavers in 2017 in Siberian Federal District which are registered in "VKontakte"; overall number is 126 thousand people. Having analyzed the most popular communities (75 thousand) among these pupils we have defined 1416 topics connected to school subjects: 73% Humanities, 18% Physics and Mathematics and 9% Natural science. Basing on this data a community classifier has been designed. The list of communities which every user out of those 126 thousand is subscribed to has been compared with the thematic classifier and the overall number has been calculated: communities of all topics, all educational communities and separately those communities which are on the Humanities, Nature Science, and Physics and Mathematics. Pupils rating on the number of subscriptions to the Humanities communities has resulted in their rating according to the interest in the Humanities.

To test the hypothesis that a pupil with multiple subscriptions to the Humanities communities has a Humanity mindset, we have studied 9390 users with the most vivid interest in the Humanities. To be considered such a user, to our mind, a pupil should have more than 75% of communities devoted to the themes of the Humanities, not less than 5 subscriptions to the Humanities communities, and not less than 5% of educational communities. Each user in this group has been suggested to agree or disagree with the idea that he or she has a Humanities mindset and plan to enter a Humanities faculty. We have received replies from 1100 pupils and 904 of them (82%) confirmed their interest in the Humanities, while 176 (18%) disagreed with that. Assuming that pupils may mistakenly evaluate

themselves those who denied their Humanities mindset have been provided with an additional test. 94 people out of those 176 have completed the diagnostic test aimed at building A.E. Golomshtok "Map of Interests". As a result, 54 respondents have shown some interest in one or several Humanities working areas at the level from +3 to +5 out of the scale from -5 to +5. Therefore, regardless respondent's denial of their inclination to the Humanities they have an interest in this field.

The study resulted in the prove of the hypothesis that analyzing pupils subscriptions to thematic communities in "VKontakte" social network one can reveal their educational interests. The accuracy of the results is 82%, while in addition to that 5% respondents show interest not only to the Technics and Nature Science, but to the Humanities, as well.

REFERENCES

1. *Schwartz H.A. et al.* Personality, gender, and age in the language of social media: The open-vocabulary approach // *PLoS one*. – 2013. – Vol. 8, № 9. – P. e73791.
2. *Kosinski M. et al.* Manifestations of user personality in website choice and behaviour on online social networks // *Machine learning*. – 2014. – Vol. 95, № 3. – P. 357–380.
3. *Goldberg L.R. et al.* The international personality item pool and the future of public-domain personality measures // *Journal of Research in personality*. – 2006. – Vol. 40, № 1. – P. 84–96.
4. *Markovikj D. et al.* Mining facebook data for predictive personality modeling // *Proceedings of the 7th international AAAI conference on Weblogs and Social Media (ICWSM 2013)*. – Boston, MA, USA, 2013.
5. *Fagerström A., Ghinea G.* Co-creation of value in higher education: using social network marketing in the recruitment of students // *Journal of Higher Education Policy and Management*. – 2013. – Vol. 35, № 1. – P. 45–53.
6. *Rutter R., Roper S., Lettice F.* Social media interaction, the university brand and recruitment performance // *Journal of Business Research*. – 2016. – Vol. 69, № 8. – P. 3096–3104.
7. *Shhebetenko S.A.* Bol'shaja pjaterka chert lichnosti i aktivnost' pol'zovatelej v social'noj seti «VKontakte» // *Vestnik Juzhno-Ural'skogo gosudarstvennogo universiteta. Serija: Psihologija*. – 2013. – T. 6, № 4.
8. *Litvinova T.A.* Vozmozhnosti komp'juternoj lingvistiki v diagnostirovanii lichnosti po tekstu // *Vestnik Voronezhskogo gosudarstvennogo universiteta. Serija: Lingvistika i mezhkul'turnaja kommunikacija*. – 2015. – № 3.
9. *Feshchenko A., Goiko V., Mozhaeva G. et al.* Analysis of user profiles in social networks to search for promising entrants // *INTED2017 Proceedings, 11th International Technology*,